Motion-based object segmentation in video sequences

ADDENDUM: VALORIZATION

Valorization is "the process of creating value from knowledge, by making knowledge suitable and/or available for social (and/or economic) use and by making knowledge suitable for translation into competitive products, services, processes and new commercial activities" (Maastricht promotie regelement, 2013). Since 2013, an addendum about valorization to the PhD dissertation is required by Maastricht University. This chapter will comply with this requirement by addressing the potential value and social relevance of the work described in this thesis. It does not form as a part of the dissertation, and should not be assessed as part of the dissertation.

Since the digital era is coming, extremely large and increasing amounts of data and informatics are produced in the economy and society. However, grasping of the useful information from a massive data is beyond the ability of humankind. In modern society, computers and machines take the work of transferring the massive, redundant digital data into high-level information that can be understand by human. To deal with visual data, such as digital images and videos, computer vision systems have been broadly and extensively investigated in both academic and industrial communities. Indeed, an ever-increasing variety of computer vision products and services in industry have been created, as the computer vision technology become more mature. From machine inspection to video surveillance, from medical image analysis to unmanned vehicles, from robots industrial to intelligent man-machine communication, etc., computer vision technology has greatly benefited the modern society. These technologies enable human to acquire useful information from videos without watching them thoroughly—by just clicking the mouse.

This thesis address an fundamental problem in computer vision and video processing: the unsupervised video object segmentation. Video object segmentation aims at extracting and analyzing object-level information from videos that can be interfaced with other thought processes and elicit appropriate action. Such techniques supplement the traditional object detection and/or recognition technology that is based on supervised machine learning. In Section 1.1, we introduced some applications of the traditional object detection and recognition technology in existing society. This thesis is innovative in respect to the existing products from the following aspects:

- This thesis solves the problem in an unsupervised way. It means that the program can automatically learn useful information from the videos. The traditional products require a large dataset of manually annotated images for training the learning model. The process of manually annotating images is fulfilled by human, which is a costly work.

- This thesis focuses on analyzing the motion information, which is a kind of feature that implied in the videos. The traditional products only analysis the visual features that are directly acquired from the videos. The motion information captures the spatial coherence of image features in a video, and provides more clues for extracting more representative information.

In conclusion, the results of this thesis addresses some short-comes of the traditional products in real-life, with potential for further development.

In general, the research results of this thesis are valuable for both academic and non-academic audiences. It can be applied in computer vision applications that requires generating and analyzing object-level information from digital videos or a sequence of images. In the following sections, we will introduce two non-academic domains that can be benefit from the research results of this thesis, as well as the representative products and/or services of these applications.

Autonomous Vehicles

Computer vision technology has been widely applied in the navigation systems for autonomous vehicles. Autonomous navigation has been extensively investigated and applied in the non-academic domain, due to its economic and social potentials. Autonomous navigation requires the vehicles or robots to perceive and understand the environment. As video data is the primary resources for autonomous navigation systems, computer vision currently plays the key role in a perception systems for autonomous vehicles or robots [125].

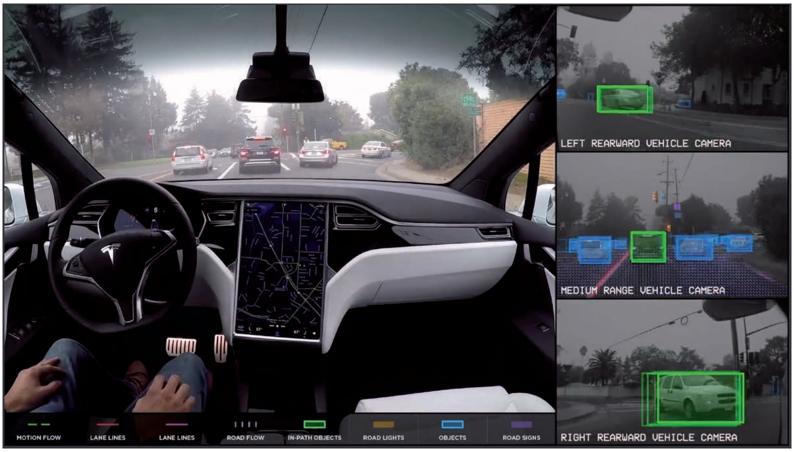

Figure 9.1: Object detection and recognition helps for understanding the environment in autonomous navigation: a snapshot of the vision system of Tesla Autopilot in driving [1].

A variety of projects has been started worldwide to explore intelligent transportation systems (ITS), by many governmental institutions and industrial groups. In Europe, the PROMETHEUS project started in 1986 was the largest project in the autonomous driving field, and defined the state of the art of autonomous vehicles [125]. Numerous universities and car manufactures, as well as research units from governments of 19 European countries participated in this project. The U.S. government established the National Automated Highway System Consortium (NAHSC) in 1995. Meanwhile Japan established the

Advanced Cruise-Assist Highway System Research Association in 1996 [25]. In the 21 century, several prototype vehicles have been developed and tested in real world, including ARGO, BRAiVE. The onboard system allows to detect obstacles, lane marking, ditches, berms and identify the presence and position of a preceding vehicle [125]. Google also started their self-driving car project in 2009. Their self-driving system equipped with different sensors, such as cameras, radars, LiDAR, wheel encoder and GPS, and can detect pedestrians, cyclists, vehicles, road work and more in all directions. Besides, many car manufactures, such as Tesla, GE, Toyota, etc., have established projects and institutes for developing autonomous driving cars. Although great progress has been made, the fully autonomous navigation cars is still in laboratory, as most of the existing computer vision systems produce errors at a rate which is not acceptable for the safety consideration. Still the achievements for autonomous vision, which refers to the computer vision technology in a autonomous navigation system, have benefited in the advanced driver assistant systems. More and more modern cars are equipped with the advanced driver assistant system. These systems help the driver in the driving process, and increase the car safety and the road safety. Some successful implementations are the Tesla Autopilot, Nissan ProPilot Assist, Mobileye, etc.

Besides the ground vehicles, like cars, computer vision system is also applied for autonomous vehicles used in different environments, such as the unmanned aerial vehicle (UAV) and autonomous underwater vehicle or robot. Vision systems for these vehicles need to deal with specific environment, and focus on specific features that are different from the ground objects as cars meet. UAVs move in 3D space and at high attitude, which requires to process videos with high resolution. UAVs have been developed and deployed many countries around the world, for applications in civilian, commercial, military, and aerospace domains. AUVs are created for various of undersea applications, such as inspection of sunken ships, sea life monitoring, military missions, undersea infrastructures or installations inspection and maintenance, etc. Vision system for AUVs deals with underwater environments, where the video quality is easily affected by the muddy or turbid waters . Currently the AUVs navigation systems are often the combination of the sonar based system and the vision based system [25].

Video Surveillance

Video surveillance systems are used to monitor security sensitive areas, such as banks, department stores, highways, crowded public places and borders. The implementation of artificial intelligence and computer vision technologies makes the video surveillance being smart: computer vision system takes over the work of human operator for tracking and detecting suspicious objects.

Video surveillance system needs to be sensitive to moving objects and providing automatic alarming function. The ability of segmentation, detection and recognition of moving objects is therefore required for an intelligent video surveillance system. Numerous of governmental institutes, universities, companies from worldwide have involved in video surveillance projects, and a variety of products and services have been produced and used in our daily life. The technologies for specific objects, such as face detection and recognition, people detection and tracking, are widely used in commercial products and offer

[1] https://www.tesla.com/autopilot

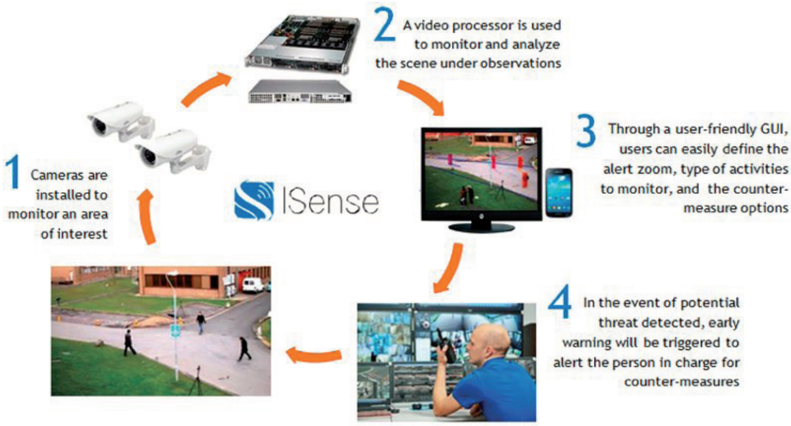

Figure 9.2: iRadar's iSense™ Smart Video Surveillance System

reliable security solutions for the society. Nowadays, numerous smart security cameras are available in markets, provide high quality intelligence security services for families with a low cost, such as the Amazon Cloud Cam, the Ring Spotlight Cam, iBaby Monitor, etc.